Appendix: Spatio-temporal Bayesian On-line Changepoint Detection with Model Selection

Jeremias Knoblauch [1] **Theodoros Damoulas** [1 2 3]

A: Condition of Theorem 1

*Denoting the spectrum of a matrix $\boldsymbol{B}$ (i.e., the set of its eigenvalues) by $\sigma(\boldsymbol{B})$, the following condition is a restatement of the relevant part in condition **A** of Meyer & Kreiss (2015)*:

Condition A. Let $\boldsymbol{W}$ be the spectral density matrix of the purely non-deterministic stochastic process $\{\boldsymbol{Y}_t\}_{t=1}^{\infty}$ satisfying the conditions of Theorem 1. We assume that the spectral density matrix is bounded, i.e. there is a constant $c > 0$ so that

$$\min\left(\sigma(\boldsymbol{W}(\lambda))\right) \geq c \tag{1}$$

for all frequencies $\lambda \in (-\pi, \pi]$, i.e. the eigenvalues of the spectral density matrix are uniformly bounded away from zero.

B: Empirical evaluation of computation time

For this comparison, we use the original code of Turner (2012) for the GP-models. As the MSE is smallest for ARGPCP for all data sets except for the snowfall data, we compare BOCPDMS against the arguably best GPCP model. We note that while NSGP performs better on the snowfall data than ARGPCP, its requirement to do Hamiltonian Monte Carlo sampling will make it significantly slower. We also note that BVAR models inside BOCPDMS outperformed the MSE of the ARGPCP model for all data sets considered. All computations were performed on a 3.1 GHz Intel i7 with 16GBRAM.

Table 1 summarizes the results. It is clear that BOCPDMS outperforms ARGPCP computationally: e.g., the computation time per parameter is between 60 (Nile data) and 585 (Bee data) times faster for BOCPDMS with BVAR models. Computation times are faster per model, too. The only exception to this is the 30 Portfolio data set, where the deployed SSBVAR models are orders of magnitude more parameter-rich than the ARGPCP-model. Related to this, we also note that comparing the computation time per parameter makes sense for two reasons: Firstly, BVARs model the d time series jointly, thus requiring d^2 parameters in the posterior covariance matrix of $\boldsymbol{y}_t$. In contrast, the GP-models ignore any dependence between the series, resulting in d parameters of the (diagonal) posterior covariance matrix for $\boldsymbol{y}$. Secondly, the parameters of the GP's kernel arguably making its parameter space Θ infinite-dimensional are not actually learnt on-line at all. Instead, they are optimized for a training period of T' observations and then fixed, see section 4 in the main paper. Hence, the parameter space the GP-models can learn in is finite-dimensional.

Table 1. Computation time in seconds per model and per parameter in the space $\Theta = \cup_{m \in \mathcal{M}} \Theta_m$

	NILE	
	TIME/$\lvert\mathcal{M}\rvert$	TIME/$\lvert\Theta\rvert$
ARGPCP	42.2	21.0
BVAR	**4.03**	**0.35**
	SNOWFALL	
	TIME/$\lvert\mathcal{M}\rvert$	TIME/$\lvert\Theta\rvert$
ARGPCP	284	142
BVAR	**157**	**4.25**
	BEE	
	TIME/$\lvert\mathcal{M}\rvert$	TIME/$\lvert\Theta\rvert$
ARGPCP	164	23.4
BVAR	**97.3**	**0.04**
	30 PORTFOLIOS	
	TIME/$\lvert\mathcal{M}\rvert$	TIME/$\lvert\Theta\rvert$
ARGPCP	**12077**	403
BVAR	34183	**1.48**

[1]Department of Statistics, University of Warwick, UK [2]Department of Computer Science, University of Warwick, UK [3]The Alan Turing Institute for Data Science & AI, UK. Correspondence to: Jeremias Knoblauch <j.knoblauch@warwick.ac.uk>.

Appendix to *Spatio-temporal Bayesian On-line Changepoint Detection with Model Selection* as published in *Proceedings of the 35^{th} International Conference on Machine Learning*, Stockholm, Sweden, PMLR 80, 2018.